MINIMUM BAYES ERROR FEATURE SELECTION

IN SPEECH RECOGNITION

Abstract of the Disclosure

In connection with speech recognition, the design of a linear transformation $\theta \in \mathfrak{R}^{p \times n}$, of rank $p \times n$, which projects the features of a classifier $\mathbf{x} \in \mathfrak{R}^n$ onto $\mathbf{y} = \theta \mathbf{x} \in \mathfrak{R}^p$ such as to achieve minimum Bayes error (or probability of misclassification). Two avenues are explored: the first is to maximize the θ-average divergence between the class densities and the second is to minimize the union Bhattacharyya bound in the range of θ. While both approaches yield similar performance in practice, they outperform standard linear discriminant analysis features and show a 10% relative improvement in the word error rate over known cepstral features on a large vocabulary telephony speech recognition task.